Mem. S.A.It. Suppl. Vol. 6, 67

Memorie della
Supplementi

Automatic classification of Phoebe's surface with the G-mode method

F. Tosi[1], A. Coradini[1,2], A.I. Gavrishin[3], and R.H. Brown[4]

[1] INAF-IFSI, Istituto di Fisica dello Spazio Interplanetario, Via del Fosso del Cavaliere, 100 I-00133 Rome, Italy, e-mail: federico.tosi@rm.iasf.cnr.it
[2] INAF-IASF, Istituto di Astrofisica Spaziale e Fisica Cosmica, Via del Fosso del Cavaliere, 100 I-00133 Rome, Italy
[3] South-Russian State Technical University, Prosveschenia 132, Novocherkassk, 346428, Russia
[4] Department of Planetary Sciences, Lunar and Planetary Lab, 1629 University Boulevard, University of Arizona, Tucson, AZ 85721, USA

Abstract. We show an example of the application of the *G-mode* method, a multivariate statistical technique for the classification of samples depending on many variables, to the infrared spectra of Phoebe acquired by VIMS (*Visual and Infrared Mapping Spectrometer*) during the Cassini flyby, occurred on 11 June 2004 while approaching Saturn. In this kind of spatially resolved observations, the G-mode is able to classify the IR calibrated spectra into several types, mainly dominated by different illumination conditions of the pixels. A phase correction for each pixel is not trivial because of the irregular shape of Phoebe. Nevertheless, the decrease of the confidence level of the G-mode test as well as the re-application of the G-mode on the main type found, lead to further types, whose differences can be in case related to superficial features. This means that an automatic spectral classification of the Saturn satellites by means of the VIMS observations is, in principle, possible. We plan to test the G-mode method also on the data coming from ongoing and future observations of the icy Saturnian satellites.

Key words. G-mode – VIMS – satellites – Phoebe – spectroscopy – classification

1. The G-mode method

The *G-mode* method was developed by A.I. Gavrishin and A. Coradini (Coradini et al. 1977) in order to explore very different kinds of classifications. Briefly, it is a clustering method which can be applied to a statistical *universe* described by N samples, each depending from M variables. It is based on the hypothesis that the frequency distribution, for each variable, follows a gaussian distribution law, provided that one studies a homogeneous group. The method can be applied without any *a priori* knowledge of the taxonomic structure of the observations (samples), and it uses informations on the *average*, the *variance* and the *correlation* among the variables, in order to find homogeneous groups of samples inside the dataset. By the reapplication of the method on these homogeneous classes, also sub-classes of statistically meaningful samples

Send offprint requests to: F. Tosi

can be recognized.
In substance, the classification criterion is based on a statistical test, and the critical value is expressed in terms of the so-called q threshold level, which defines the confidence level of the test: the larger is q, the more general is the classification for given errors on the variables. By changing this value it is therefore possible to get different levels of classification and correspondingly different degrees of class homogeneity. Vice versa, for a fixed q value, the classification becomes more general increasing the errors on the variables.

2. The VIMS data set

VIMS is an imaging spectrometer onboard the Cassini orbiter. VIMS is, in turn, made up of two spectrometers, VIMS-V and VIMS-IR, designed for the visible and infrared spectral range, respectively (see Brown et al. (2004a) for more details). The VIMS data are characterized by two spatial dimensions and one spectral dimension. In pratice, this means that each pixel inside the image is sampled in 352 spectral bands ranging from 0.35 μm to 5.1 μm (96 bands for VIMS-V, 256 bands for VIMS-IR). Therefore, the final result of every acquisition is an "image cube" (see Fig. 1). However, in this work we considered only the IR portion of the VIMS cubes, because, generally, IR spectra can discriminate the composition of solid surfaces, showing many diagnostic features. A more detailed treatment can be found in Tosi et al. (2005); here we summarize the main results of this application.

In this case, our statistical universe is composed by N spectra (each corresponding to a pixel of the satellite inside the IR image) sampled in M variables (the 256 VIMS-IR spectral channels ranging from 0.85 μm to 5.1 μm, so in our case M = 256). Errors on the variables can be set both as absolute (i.e., equal for all the variables) or relative (i.e., every variable has its own error, given by the instrumental noise of the corresponding spectral band).
For our purpose, the satellite image must be spatially resolved. Among the 288 acquisitions by VIMS during the Phoebe flyby, we selected one cube. This is the cube coded V1465679932_1, acquired at 20:54 (UTC onboard the spacecraft) with 380 msec IR integration time, during the subsequence PHOEBE053, with the spacecraft about 30,559 km from the satellite (solar phase angle of 89.2 degrees). In this cube, with spatial dimensions of 48×48 pixels, we have a medium spatial resolution (≈15 km/pixel), as well as some sky background pixels that are useful to perform a valid background subtraction (see Fig. 2).

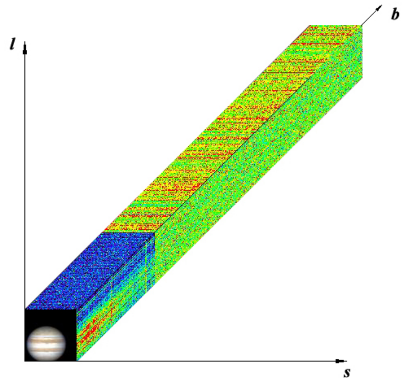

Fig. 1. Structure of the VIMS data. The *l* and *s* axes are the spatial dimensions, while *b* is the spectral dimension.

In this IR image, we selected 158 pixels of the satellite with a sufficiently high raw signal. Thus the dataset consists of 158 infrared spectra of Phoebe acquired by VIMS-IR, each sampled in 256 wavelengths between 0.85 and 5.1 μm. Since we want to attempt an automatic mapping of the image, we keep the information about the coordinates of every pixel inside of the image.
All of the raw spectra were calibrated into I/F (reflectance) spectra by means of the latest VIMS-IR sensitivity function, and despiked, as well. No phase correction was applied to the spectra, because, given the irregular shape of Phoebe, every correction would be not trivial. This has great importance, because it lets us foresee that most of the spectra will be domi-

Fig. 2. Image of Phoebe in the IR portion (band 121, λ_{eff}=1279.7 nm) of the VIMS cube V1465679932_1.

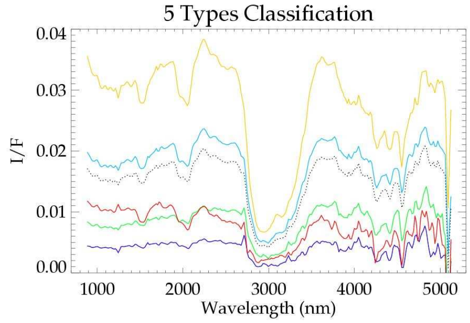

Fig. 3. Average I/F spectra of the 5 types isolated by the G-mode, using an absolute error of 0.001 and q = 2.43. The spectra are shown without error bars for a better comparison. The black dotted line represents the global average of all the samples.

nated by the location and phase of the individual pixels to which they correspond.

3. Application of the G-mode

Because of Phoebe's low albedo (maximum I/F is ≈0.05 in the IR portion of the selected cube), and its consequent low SNR, we initially assumed the same error for all of the 256 variables, so that the statistical weights of the variables become of the same order of magnitude. We first applied an absolute error of 0.001, ten times smaller than Phoebe's typical reflectance, in order to take into account the natural fluctuation of the variables around their average value. By assuming q = 2.43, corresponding to 99.25% confidence level, the classification leads to 5 types. The average spectra (i.e., mean values of the variables) representing these types are revealed in Fig. 3.

Even if this work was not intended to interpret the spectral features (for this purpose, see Clark et al. (2005)), two results can be easily inferred. First, as expected, the types found are mostly dominated by the different reflectance of the samples (pixels); secondly, the classification permits to discriminate some different features, which evidently tend to disappear in the global average. Through the information on the spatial position of the spectra, it is possible to reconstruct a spectral map of the image (see Fig. 4).

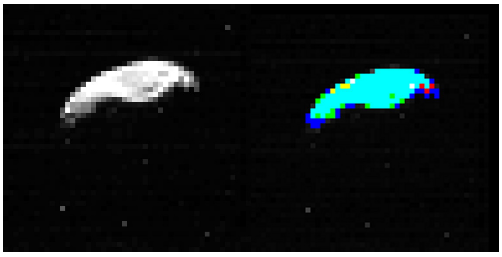

Fig. 4. Spectral map of Phoebe derived by the G-mode method, using an absolute error of 0.001 and threshold q=2.43. On the left is the original image. The colours are related to the spectra shown in Fig. 3. The colourless pixels are relative to spectra discarded by the processing.

By lowering the q value, that is, with a lesser confidence level, more types can be found, as expected. However, these types appear again to be dominated by a "phase effect", given by the different phase of each pixel. Then, after the first classification, which was performed with an absolute error for all of the variables, we can consider the results obtained with a relatively high threshold value, and re-apply the G-mode only on the largest type (that is, the first type, containing 115 samples), using a lower q value in order to reduce the phase effect. To do this, there are two possible procedures: 1) keeping the errors on the variables at an absolute level of 0.001, as before, or 2) normalizing the spectra at one specific spectral channel and assigning to the variables their "real" errors (for each channel, this is given by the inverse of its SNR).

In the first procedure, that is absolute errors without normalization, by applying q=1.75 (96% (confidence level), we find 7 subtypes, whose average spectra are summarized in Fig. 5 and mapped on the IR image in Fig. 6.

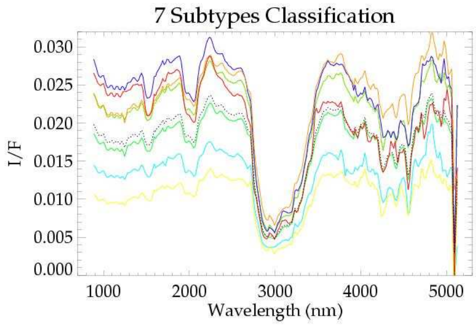

Fig. 5. Summary of the 7 subtypes found reapplying the G-mode on the first type (found with the classification with absolute errors). The spectra are shown without error bars for a better comparison. The black dotted line represents the average of the 115 samples belonging to the original type.

In the second approach, all of the spectra were normalized to the spectral channel #179, corresponding to λ_{eff} = 2.234 μm (this band is suitable in order to sample the continuum, and does not coincide with a spectral feature). In this case, the statistical weights of the variables encompass more orders of magnitude, so that the classification is performed on the basis of a subset of variables. As a result, for all of the explored q values, the G-mode finds only 2 or 3 subtypes. Anyway, another interesting application was attempted in order to test the ability of the method to find different types using a lower number of variables. We reapplied the G-mode on the first type originally found, considering

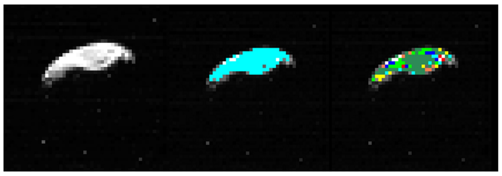

Fig. 6. Spectral map derived by the re-application of the G-mode on the first type, again using an absolute error of 0.001. On the left is the original IR image, in the middle the first type (cyan) is pointed out, and finally on the right is the re-classification of this type, that, with a threshold value q=1.75, appears in turn divided into 7 subtypes. The colours are related to the spectra shown in Fig. 5.

only the spectral range from 3.5 μm to 5.1 μm. In this case, we have 96 variables (the last 96 spectral channels of VIMS-IR), instead of 256. The selected spectral range is particularly suitable for this test, because it shows some interesting absorption features, while, at the same time, it is not affected by the large water ice absorption feature centred at ≈3 μm.

The result is that the classification appears to be more sensitive to the q threshold value: more subtypes are found, even though only one of these appears to have a large number of samples. For instance, by assuming q=2.58, which corresponds to 99.5% confidence level, we find 6 subtypes, only one of which (subtype #4) has 94 samples (see Fig. 7 and Fig. 8).

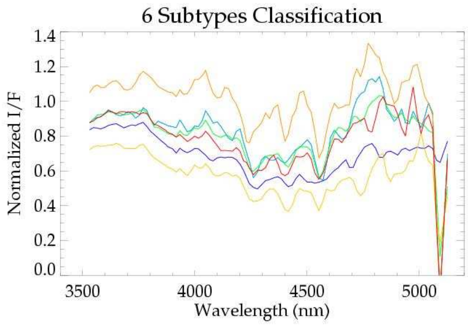

Fig. 7. Summary of the 6 subtypes found reapplying the G-mode on the first type (came out from the first classification with absolute errors). In this case, only the 3.5 - 5.1 μm region (96 variables) is considered, using normalized spectra with relative errors and q=2.58.

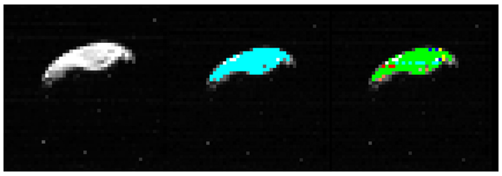

Fig. 8. On the right: a spectral map derived by the re-application of the G-mode on the first type (cyan) in the 3.5 - 5.1 μm range, using normalized spectra with relative errors and q=2.58. The colours are related to the spectra shown in Fig. 7.

4. Conclusions

Applied to the resolved VIMS observations of Phoebe, the G-mode is able to separate some "types" on the surface of the satellite. By lowering the confidence level of the statistical test, more types can be found, but, because of the satellite's irregular shape, which does not permit an easy phase correction, they are mainly related to different illumination conditions of the individual pixels. Nonetheless, within these classes, spectral differences can be recognized, as well. These differences can be emphasized by the reapplication of the method on the largest type found by the first classification. Since this can be performed through different procedures, we generally found that the method is very sensitive to the errors applied on the variables, rather than to the confidence level of the test. This seems to be due to the fact that an absolute error gives more statistical weight to a larger number of variables, while with pure instrumental errors the classification is essentially made on the basis of a lesser number of variables.

References

Brown, R. H., et al. 2004a, Space Sci. Rev., in press

Coradini, A., et al. 1977, Comput. Geosci. 3, 85-105

Clark, R. N., et al. 2005, Nature, in press

Tosi, F., et al. 2005, submitted to Earth, Moon, and Planets. Under review.